Comparison Tables: BBOB 2015 Testbed in 5-D (Expensive Setting)

The BBOBies

July 16, 2015

Abstract

This document provides tabular results of the workshop on Black-Box Optimization Benchmarking held at GECCO 2015 with a focus on benchmarking black-box algorithms for small function evaluation budgets ("expensive setting"), see `http://coco.gforge.inria.fr/doku.php?id=bbob-2015`. Overall, 18 algorithms have been tested on 24 benchmark functions in dimensions between 2 and 20. Only three of them have been tested on the optional instances in dimension 40. A description of the used objective functions can be found in [7, 5]. The experimental set-up is described in [6].

The performance measure provided in the following tables is the expected number of objective function evaluations to reach a given target function value (ERT, expected running time), divided by the respective value for the best algorithm in BBOB-2009 (see [2]) if an algorithm from BBOB-2009 reached the given target function value. The ERT value is given otherwise ($\mathrm{ERT}_{\mathrm{best}}$ is noted as infinite). See [6] for details on how ERT is obtained. Bold entries in the table correspond to values below 3 or the top-three best values. Table 1 gives an overview on all algorithms submitted to the noise-free testbed at GECCO 2015.

Table 1: Names and references of all algorithms submitted for the noise-free testbed

algorithm short name	paper	reference
BSifeg	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
BSif	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
BSqi	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
BSrr	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
CMA-CSA	Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB Noiseless Testbed	[1]
CMA-MSR	Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB Noiseless Testbed	[1]
CMA-TPA	Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB Noiseless Testbed	[1]
GP1-CMAES	SBenchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
GP5-CMAES	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
IPOPCMAv3p61	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
LHD-10xDefault-MATSuMoT	The Impact of Initial Designs on the Performance of MATSuMoTo on the Noiseless BBOB-2015 Testbed: A Preliminary Study	[4]
LHD-2xDefault-MATSuMoTo	The Impact of Initial Designs on the Performance of MATSuMoTo on the Noiseless BBOB-2015 Testbed: A Preliminary Study	[4]
RAND-2xDefault-MATSuMoTo	The Impact of Initial Designs on the Performance of MATSuMoTo on the Noiseless BBOB-2015 Testbed: A Preliminary Study	[4]
RF1-CMAES	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
RF5-CMAES	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
Sifeg	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
Sif	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
Srr	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]

Table 2: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_1 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f1	*2.5e+1*:4.8	*1.6e+1*:7.6	*1.0e-8*:12	*1.0e-8*:12	*1.0e-8*:12	15/15
BSifeg	**1.8**(2)	**1.7**(0.8)	**2.2**(0.1)	**2.2**(0.1)	**2.2**(0.1)	15/15
BSif	**1.8**(2)	**1.7**(1)	**2.2**(0.2)	**2.2**(0.1)	**2.2**(0.2)	15/15
BSqi	**1.8**(2)	**1.7**(1)	**2.2**(0.1)	**2.2**(0.1)	**2.2**(0.2)	15/15
BSrr	**1.8**(2)	**1.7**(0.9)	**2.2**(0.0)	**2.2**(0.2)	**2.2**(0.2)	15/15
CMA-CSA	4.3(3)	3.6(2)	58(5)	58(2)	58(5)	15/15
CMA-MSR	3.0(2)	3.1(3)	90(7)	90(9)	90(9)	15/15
CMA-TPA	**2.7**(4)	**2.5**(3)	52(9)	52(4)	52(10)	15/15
GP1-CMAES	**2.5**(2)	**2.2**(2)	36(6)	36(5)	36(5)	15/15
GP5-CMAES	**2.3**(2)	**2.1**(1)	92(54)	92(86)	92(122)	11/15
IPOPCMAv3p	3.0(4)	**2.9**(4)	59(5)	59(9)	59(7)	15/15
LHD-10xDef	**2.8**(1)	3.8(6)	∞	∞	∞ *250*	0/15
LHD-2xDefa	**2.6**(2)	**2.8**(1)	∞	∞	∞ *250*	0/15
RAND-2xDef	**2.0**(1)	**2.2**(1)	∞	∞	∞ *250*	0/15
RF1-CMAES	3.7(3)	**3.0**(1)	1520(1031)	1520(2191)	1520(799)	1/15
RF5-CMAES	**2.2**(3)	**2.4**(1)	∞	∞	∞ *1252*	0/15
Sifeg	**1.8**(2)	**1.7**(1)	8.9(0.9)	8.9(1.0)	8.9(1)	15/15
Sif	**1.8**(2)	**1.7**(0.7)	8.7(1)	8.7(1)	8.7(0.9)	15/15
Srr	**1.8**(1)	**1.7**(1)	8.4(1)	8.4(1)	8.4(1)	15/15

Table 3: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_2 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f2	*1.6e+6*:2.9	*4.0e+5*:11	*4.0e+4*:15	*6.3e+2*:58	*1.0e-8*:95	15/15
BSifeg	6.0(3)	**1.8**(1)	**1.5**(0.2)	**0.52**(0.2)	**1.1**(0.1)	15/15
BSif	6.0(4)	**1.8**(1)	**1.5**(0.2)	**0.53**(0.3)	**1.1**(0.2)	15/15
BSqi	6.0(4)	**1.8**(1)	**1.5**(0.2)	**0.49**(0.1)	**0.92**(0.2)	15/15
BSrr	6.0(2)	**1.8**(1)	**1.5**(0.2)	**0.52**(0.2)	**1.1**(0.2)	15/15
CMA-CSA	**2.3**(2)	**1.8**(2)	6.1(4)	7.5(2)	18(0.9)	15/15
CMA-MSR	**2.5**(2)	**1.9**(2)	5.9(4)	7.4(3)	21(1)	15/15
CMA-TPA	3.7(6)	**2.1**(2)	3.9(3)	7.7(3)	18(3)	15/15
GP1-CMAES	**2.1**(1.0)	**1.9**(2)	4.0(3)	5.8(4)	∞ *1258*	0/15
GP5-CMAES	**2.5**(3)	**1.7**(2)	3.2(1)	**2.7**(1)	94(92)	2/15
IPOPCMAv3p	3.1(4)	**1.5**(2)	5.6(3)	10(9)	∞ *1258*	0/15
LHD-10xDef	**1.1**(0.7)	**1.2**(2)	8.4(6)	65(37)	∞ *250*	0/15
LHD-2xDefa	**1.3**(1)	**0.79**(2)	4.3(2)	32(47)	∞ *250*	0/15
RAND-2xDef	**1.6**(1)	**1.0**(0.6)	3.3(2)	62(33)	∞ *250*	0/15
RF1-CMAES	**2.9**(4)	**2.7**(2)	6.8(5)	51(50)	∞ *1258*	0/15
RF5-CMAES	**2.1**(1)	7.1(23)	25(29)	∞	∞ *1260*	0/15
Sifeg	6.4(4)	**2.0**(1)	**2.1**(0.6)	**0.84**(0.2)	**1.4**(0.1)	15/15
Sif	6.4(2)	**2.0**(0.2)	**2.1**(0.6)	**0.82**(0.1)	**1.4**(0.3)	15/15
Srr	6.4(3)	**2.0**(0.6)	**2.1**(0.5)	**0.79**(0.1)	**1.5**(0.2)	15/15

Table 4: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_3 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f3	*1.6e+2*:4.1	*1.0e+2*:15	*6.3e+1*:23	*2.5e+1*:73	*1.0e+1*:716	15/15
BSifeg	**2.6**(2)	**1.2**(0.2)	**0.98**(0.1)	**0.46**(0.1)	**0.11**(0.1)	15/15
BSif	**2.6**(2)	**1.2**(0.2)	**0.98**(0.1)	**0.46**(0.1)	**0.11**(0.0)	15/15
BSqi	**2.6**(2)	**1.2**(0.5)	**0.98**(0.1)	**0.46**(0.1)	**0.10**(0.1)	15/15
BSrr	**2.6**(2)	**1.2**(0.4)	**0.98**(0.1)	**0.46**(0.1)	**0.09**(0.1)	15/15
CMA-CSA	4.2(5)	**2.4**(2)	**2.6**(2)	**2.5**(1)	**1.4**(0.3)	15/15
CMA-MSR	3.3(3)	**2.4**(1)	3.4(2)	6.7(14)	**1.7**(2)	15/15
CMA-TPA	3.5(7)	**2.2**(1)	**2.6**(2)	**2.5**(1)	**0.81**(0.7)	15/15
GP1-CMAES	**2.8**(3)	**1.5**(2)	**2.3**(0.7)	**2.5**(5)	**1.6**(1)	11/15
GP5-CMAES	3.0(2)	**1.3**(0.9)	**1.6**(1)	**2.2**(0.5)	**2.6**(3)	8/15
IPOPCMAv3p	**2.7**(2)	**1.4**(2)	**2.5**(0.8)	3.1(1)	**1.1**(0.9)	12/15
LHD-10xDef	**2.1**(3)	**2.8**(2)	3.7(2)	3.6(0.5)	**1.0**(0.8)	5/15
LHD-2xDefa	**2.1**(2)	**1.5**(1)	**1.9**(1)	**2.1**(3)	**2.5**(4)	2/15
RAND-2xDef	**2.3**(2)	**1.4**(1)	**1.6**(0.4)	**2.2**(0.5)	**0.58**(0.3)	7/15
RF1-CMAES	**1.9**(3)	**1.6**(1)	**2.2**(0.8)	**2.4**(1)	3.0(3)	6/15
RF5-CMAES	3.4(3)	**1.6**(1)	4.6(19)	14(15)	6.1(7)	4/15
Sifeg	**2.6**(2)	**1.2**(0.5)	**0.99**(0.1)	**0.57**(0.4)	**0.13**(0.1)	15/15
Sif	**2.6**(2)	**1.2**(0.3)	**0.99**(0.1)	**0.57**(0.2)	**0.13**(0.1)	15/15
Srr	**2.6**(2)	**1.2**(0.3)	**0.99**(0.2)	**0.56**(0.1)	**0.12**(0.0)	15/15

Table 5: 05-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_4 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f4	*2.5e+2*:2.6	*1.6e+2*:10	*1.0e+2*:19	*4.0e+1*:65	*1.6e+1*:434	15/15
BSifeg	**2.1**(2)	**0.90**(0.8)	**1.4**(0.8)	**1.0**(0.6)	**0.25**(0.1)	15/15
BSif	**2.1**(3)	**0.90**(0.7)	**1.4**(0.5)	**1.1**(0.5)	**0.26**(0.1)	15/15
BSqi	**2.1**(2)	**0.90**(1)	**1.3**(0.4)	**1.3**(0.8)	**0.28**(0.2)	15/15
BSrr	**2.1**(2)	**0.90**(0.9)	**1.3**(1.0)	**1.1**(0.4)	**0.23**(0.1)	15/15
CMA-CSA	4.6(9)	**2.9**(4)	3.0(2)	4.3(1)	**2.3**(3)	15/15
CMA-MSR	4.2(6)	**2.7**(4)	**2.3**(1)	**2.8**(2)	**2.0**(3)	15/15
CMA-TPA	4.7(5)	**2.5**(1)	3.4(2)	**3.0**(2)	**2.4**(2)	15/15
GP1-CMAES	**2.7**(2)	**2.4**(2)	**2.6**(3)	5.1(8)	3.2(2)	9/15
GP5-CMAES	**1.6**(1)	**1.6**(1)	**3.0**(2)	5.2(6)	7.5(7)	5/15
IPOPCMAv3p	**2.3**(3)	**1.5**(1)	**2.4**(1)	**2.4**(1)	**1.2**(0.9)	14/15
LHD-10xDef	3.9(2)	3.2(3)	5.1(3)	8.3(6)	∞ *250*	0/15
LHD-2xDefa	**2.2**(3)	**1.5**(1)	3.0(2)	6.8(11)	∞ *250*	0/15
RAND-2xDef	**2.8**(3)	**2.2**(2)	3.1(2)	4.2(3)	8.4(13)	1/15
RF1-CMAES	**2.4**(3)	**1.9**(0.8)	**2.5**(2)	7.9(10)	12(20)	3/15
RF5-CMAES	**2.7**(4)	**1.9**(1)	4.9(11)	31(26)	∞ *1252*	0/15
Sifeg	**2.1**(2)	**0.92**(1.0)	**1.3**(0.4)	**0.85**(0.3)	**0.21**(0.1)	15/15
Sif	**2.1**(3)	**0.92**(0.7)	**1.3**(0.5)	**0.85**(0.3)	**0.21**(0.1)	15/15
Srr	**2.1**(2)	**0.92**(1)	**1.3**(0.8)	**0.83**(0.2)	**0.20**(0.0)	15/15

Table 6: 05-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_5 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f5	*6.3e+1*:4.0	*4.0e+1*:10	*1.0e-8*:10	*1.0e-8*:10	*1.0e-8*:10	15/15
BSifeg	**2.9**(0.7)	**1.4**(0.2)	**1.5**(0.1)	**1.5**(0.1)	**1.5**(0.1)	15/15
BSif	**2.9**(0.4)	**1.4**(0.2)	**1.5**(0.1)	**1.5**(0.1)	**1.5**(0.1)	15/15
BSqi	**2.9**(0.9)	**1.4**(0.2)	**1.5**(0.1)	**1.5**(0.1)	**1.5**(0.1)	15/15
BSrr	**2.9**(0.4)	**1.4**(0.2)	**1.5**(0.1)	**1.5**(0.1)	**1.5**(0.1)	15/15
CMA-CSA	**1.9**(2)	**1.8**(0.7)	5.2(2)	5.2(2)	5.2(2)	15/15
CMA-MSR	**2.2**(2)	**1.8**(1)	5.9(3)	5.9(2)	5.9(1)	15/15
CMA-TPA	**2.6**(3)	**1.9**(2)	5.1(2)	5.1(2)	5.1(2)	15/15
GP1-CMAES	**2.0**(2)	**1.5**(1.0)	26(9)	26(20)	26(37)	15/15
GP5-CMAES	3.0(1)	**1.7**(0.1)	6.4(4)	6.4(4)	6.4(3)	15/15
IPOPCMAv3p	3.0(3)	**2.2**(1)	21(12)	21(20)	21(14)	15/15
LHD-10xDef	3.2(5)	4.6(5)	13(0.2)	13(0.2)	13(0.2)	15/15
LHD-2xDefa	**1.9**(1)	**1.8**(1)	3.5(2)	3.5(2)	3.5(3)	15/15
RAND-2xDef	**2.0**(2)	**2.0**(1.0)	3.1(0.2)	3.1(0.2)	3.1(0.2)	15/15
RF1-CMAES	**2.3**(2)	**1.7**(0.9)	45(26)	45(35)	45(26)	15/15
RF5-CMAES	**2.8**(3)	**1.9**(2)	137(220)	137(56)	137(117)	10/15
Sifeg	**2.9**(0.8)	**1.4**(0.2)	**1.5**(0.1)	**1.5**(0.1)	**1.5**(0.1)	15/15
Sif	**2.9**(0.3)	**1.4**(0.1)	**1.5**(0.1)	**1.5**(0.1)	**1.5**(0.1)	15/15
Srr	**2.9**(0.8)	**1.4**(0.2)	**1.5**(0.1)	**1.5**(0.1)	**1.5**(0.1)	15/15

Table 7: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_6 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f6	*1.0e+5*:3.0	*2.5e+4*:8.4	*1.0e+2*:16	*2.5e+1*:54	*2.5e-1*:254	15/15
BSifeg	**1.2**(0.9)	**1.2**(0.8)	**2.3**(3)	66(371)	210(215)	9/15
BSif	**1.2**(1)	**1.2**(1.0)	**2.9**(6)	79(83)	1315(1887)	2/15
BSqi	**1.2**(2)	**1.2**(0.8)	3.6(0.8)	71(214)	228(417)	8/15
BSrr	**1.2**(2)	**1.2**(1)	**2.3**(1)	66(403)	226(149)	8/15
CMA-CSA	**2.0**(2)	**1.6**(0.9)	**2.9**(1)	**2.7**(2)	**1.9**(0.5)	15/15
CMA-MSR	3.2(3)	**1.9**(1)	3.5(4)	3.3(1)	**2.1**(0.5)	15/15
CMA-TPA	**2.9**(2)	**2.0**(2)	5.9(3)	3.5(2)	**2.0**(0.2)	15/15
GP1-CMAES	**2.2**(3)	**1.4**(1)	**2.3**(2)	**2.2**(2)	74(77)	1/15
GP5-CMAES	**3.0**(3)	**1.8**(1)	**2.8**(7)	3.7(7)	∞ *1260*	0/15
IPOPCMAv3p	3.1(4)	**1.8**(1)	3.3(3)	**3.1**(3)	**2.2**(0.6)	15/15
LHD-10xDef	**1.6**(2)	**2.2**(4)	5.5(4)	3.6(3)	∞ *250*	0/15
LHD-2xDefa	**1.7**(2)	**1.3**(1)	3.2(5)	4.9(5)	∞ *250*	0/15
RAND-2xDef	**2.3**(2)	**1.8**(1)	**2.8**(4)	9.0(8)	∞ *250*	0/15
RF1-CMAES	**3.0**(3)	**2.0**(2)	4.5(5)	8.5(4)	∞ *1258*	0/15
RF5-CMAES	3.6(4)	**2.4**(3)	19(54)	154(117)	∞ *1260*	0/15
Sifeg	**1.2**(1)	**1.2**(0.9)	5.3(0.8)	36(10)	120(143)	12/15
Sif	**1.2**(2)	**1.2**(1.0)	8.2(26)	64(388)	458(237)	5/15
Srr	**1.2**(1)	**1.2**(1)	6.7(21)	34(112)	101(59)	11/15

Table 8: 05-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_7 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f7	*1.6e+2*:4.2	*1.0e+2*:6.2	*2.5e+1*:20	*4.0e+0*:54	*1.0e+0*:324	15/15
BSifeg	34(249)	24(2)	55(128)	620(1442)	754(1633)	3/15
BSif	34(2)	24(85)	64(137)	600(462)	1037(1119)	2/15
BSqi	34(1)	24(2)	68(76)	488(365)	726(522)	3/15
BSrr	35(127)	46(168)	96(0.6)	669(912)	1050(649)	2/15
CMA-CSA	**2.0**(2)	**2.8**(1)	3.6(3)	3.9(1.0)	**1.3**(1.0)	15/15
CMA-MSR	3.5(4)	3.7(3)	3.2(3)	4.5(8)	**1.1**(1.0)	15/15
CMA-TPA	**2.5**(3)	**2.3**(2)	**2.5**(2)	**2.7**(1)	**0.98**(0.1)	15/15
GP1-CMAES	**1.7**(1)	**2.0**(1)	**2.1**(1)	4.7(2)	**1.4**(1.0)	15/15
GP5-CMAES	**2.2**(2)	**2.5**(2)	**1.9**(0.9)	**1.2**(0.6)*	**0.82**(0.9)	15/15
IPOPCMAv3p	3.3(2)	3.4(4)	3.3(2)	**3.5**(0.9)	**1.5**(0.7)	14/15
LHD-10xDef	**1.5**(1)	**2.2**(3)	5.2(2)	4.8(3)	5.5(7)	2/15
LHD-2xDefa	**1.4**(1)	**2.0**(3)	**1.8**(2)	8.0(8)	11(11)	1/15
RAND-2xDef	**1.3**(1.0)	**1.7**(2)	**2.5**(3)	9.3(7)	11(24)	1/15
RF1-CMAES	**1.9**(2)	**2.6**(1)	**2.7**(2)	15(15)	10(9)	5/15
RF5-CMAES	**2.5**(3)	**2.7**(2)	5.3(13)	33(23)	17(16)	3/15
Sifeg	**1.6**(3)	**2.0**(3)	**1.5**(2)	396(707)	276(160)	6/15
Sif	**1.6**(2)	**2.0**(2)	8.1(26)	296(420)	204(167)	8/15
Srr	**1.6**(3)	**2.1**(2)	8.3(27)	317(429)	306(449)	6/15

Table 9: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_8 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f8	*1.0e+4*:4.6	*6.3e+3*:6.8	*1.0e+3*:18	*6.3e+1*:54	*1.6e+0*:258	15/15
BSifeg	**1.6**(0.3)	**1.8**(1)	**1.0**(0.1)	13(30)	38(40)	15/15
BSif	**1.6**(2)	**1.8**(1)	**1.0**(0.2)	34(73)	64(66)	14/15
BSqi	**1.6**(2)	**1.8**(1)	**1.0**(0.4)	8.2(11)	35(49)	14/15
BSrr	**1.6**(2)	**1.8**(1)	**1.0**(0.1)	12(12)	37(37)	14/15
CMA-CSA	**2.6**(1)	**2.8**(1)	**2.8**(1)	**2.7**(1)	**4.9**(1)	15/15
CMA-MSR	**2.2**(3)	**2.0**(4)	**2.6**(3)	4.2(1)	**3.4**(2)	15/15
CMA-TPA	3.4(6)	**2.6**(6)	**2.7**(2)	4.0(1)	5.8(5)	15/15
GP1-CMAES	**2.8**(3)	**2.2**(3)	**2.3**(1)	**2.1**(0.9)	7.2(11)	8/15
GP5-CMAES	**2.1**(1)	**1.9**(2)	**1.9**(1)	**2.5**(2)	70(35)	1/15
IPOPCMAv3p	3.1(3)	**2.8**(2)	**2.1**(1.0)	3.8(2)	**4.9**(4)	11/15
LHD-10xDef	**2.7**(3)	**2.8**(2)	5.2(3)	12(8)	∞ *250*	0/15
LHD-2xDefa	**2.4**(1)	**2.3**(2)	**1.5**(0.9)	3.6(9)	∞ *250*	0/15
RAND-2xDef	3.1(2)	**2.3**(2)	**1.5**(0.6)	3.1(2)	∞ *250*	0/15
RF1-CMAES	**2.6**(2)	**2.3**(2)	3.1(2)	7.8(12)	36(38)	2/15
RF5-CMAES	**2.3**(2)	**2.0**(1)	**2.8**(3)	56(29)	∞ *1252*	0/15
Sifeg	**1.6**(2)	**1.8**(1)	**1.1**(0.1)	**2.2**(3)	17(22)	15/15
Sif	**1.6**(2)	**1.8**(1)	**1.1**(0.5)	3.4(1)	48(43)	14/15
Srr	**1.6**(1)	**1.8**(1)	**1.1**(0.1)	**1.7**(1)	21(34)	15/15

Table 10: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_9 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f9	*2.5e+1*:20	*1.6e+1*:26	*1.0e+1*:35	*4.0e+0*:62	*1.6e-2*:256	15/15
BSifeg	19(58)	16(75)	14(16)	223(484)	∞ *5e4*	0/15
BSif	57(363)	45(279)	36(15)	416(392)	∞ *5e4*	0/15
BSqi	15(6)	13(0.9)	11(28)	181(58)	∞ *5e4*	0/15
BSrr	17(13)	14(20)	15(14)	248(306)	∞ *4e4*	0/15
CMA-CSA	**7.5**(2)	**6.7**(2)	**5.7**(0.6)	**6.6**(7)	**7.1**(4)	15/15
CMA-MSR	10(4)	8.5(2)	7.2(1)	7.9(3)	**6.9**(7)	15/15
CMA-TPA	7.8(3)	**6.8**(3)	**5.4**(2)	**4.6**(2)	**5.0**(1)	15/15
GP1-CMAES	12(10)	10(15)	8.2(7)	14(24)	∞ *1258*	0/15
GP5-CMAES	15(5)	14(12)	13(19)	27(17)	70(55)	1/15
IPOPCMAv3p	10(3)	8.7(3)	7.5(2)	**7.2**(10)	36(18)	2/15
LHD-10xDef	185(180)	144(128)	∞	∞	∞ *250*	0/15
LHD-2xDefa	12(3)	12(7)	25(40)	∞	∞ *250*	0/15
RAND-2xDef	11(5)	11(9)	20(3)	60(85)	∞ *250*	0/15
RF1-CMAES	38(82)	36(51)	30(15)	48(42)	∞ *1258*	0/15
RF5-CMAES	106(204)	126(50)	257(205)	∞	∞ *1252*	0/15
Sifeg	**7.5**(6)	6.9(14)	5.8(4)	102(220)	∞ *5e4*	0/15
Sif	36(105)	28(93)	24(4)	221(505)	∞ *5e4*	0/15
Srr	**6.2**(2)	**5.3**(6)	**4.5**(7)	95(237)	∞ *4e4*	0/15

Table 11: 05-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{10} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f10	*2.5e+6*:2.9	*6.3e+5*:7.0	*2.5e+5*:17	*6.3e+3*:54	*2.5e+1*:297	15/15
BSifeg	**1.8**(2)	**2.1**(0.9)	12(41)	463(716)	∞ *3e4*	0/15
BSif	**1.8**(2)	**2.1**(1)	9.2(0.7)	481(550)	∞ *3e4*	0/15
BSqi	**1.8**(2)	**2.1**(1)	7.1(0.4)	325(448)	∞ *3e4*	0/15
BSrr	**1.8**(1)	**2.1**(1)	10(0.6)	529(749)	∞ *2e4*	0/15
CMA-CSA	**2.5**(2)	**1.8**(1)	**1.5**(0.8)	4.0(1)	**2.7**(0.6)	15/15
CMA-MSR	**2.3**(1)	**2.0**(3)	**1.6**(1)	4.0(3)	**2.9**(0.8)	15/15
CMA-TPA	**1.9**(3)	**2.3**(2)	**1.4**(2)	**3.3**(2)	**2.3**(1)	15/15
GP1-CMAES	**1.2**(2)	**0.94**(1.0)	**0.85**(0.6)	**2.9**(2)	**2.0**(0.4)	15/15
GP5-CMAES	**2.4**(3)	**1.6**(1)	**1.0**(1)	**1.6**(0.6)	**0.95**(0.2)	15/15
IPOPCMAv3p	**1.8**(1)	**2.2**(2)	**1.6**(2)	4.7(2)	4.4(5)	11/15
LHD-10xDef	**1.5**(2)	**1.7**(0.9)	**1.4**(2)	12(9)	∞ *250*	0/15
LHD-2xDefa	**1.9**(2)	**1.4**(1)	**1.6**(1)	12(13)	∞ *250*	0/15
RAND-2xDef	**2.3**(2)	**2.1**(2)	**1.5**(0.9)	3.5(1)	∞ *250*	0/15
RF1-CMAES	3.2(4)	**2.3**(2)	**1.6**(0.8)	10(18)	63(90)	1/15
RF5-CMAES	**2.7**(2)	**1.7**(2)	4.5(0.4)	43(28)	∞ *1260*	0/15
Sifeg	**1.8**(2)	**1.9**(1)	**1.5**(1)	76(134)	∞ *1e4*	0/15
Sif	**1.8**(2)	**1.9**(2)	**1.4**(1)	97(142)	∞ *1e4*	0/15
Srr	**1.8**(2)	**1.9**(1)	**1.4**(1)	93(70)	∞ *1e4*	0/15

Table 12: 05-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{11} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f11	*1.0e+6*:3.0	*6.3e+4*:6.2	*6.3e+2*:16	*6.3e+1*:74	*6.3e-1*:298	15/15
BSifeg	**2.4**(3)	**2.2**(1)	**1.5**(0.5)	116(123)	∞ *3e4*	0/15
BSif	**2.4**(1)	**2.2**(1)	**1.5**(0.5)	93(86)	∞ *3e4*	0/15
BSqi	**2.4**(2)	**2.2**(1)	**1.5**(0.6)	64(103)	∞ *4e4*	0/15
BSrr	**2.4**(1)	**2.2**(1)	**1.5**(0.6)	85(79)	∞ *3e4*	0/15
CMA-CSA	**1.5**(1)	**2.0**(2)	3.8(2)	6.0(4)	**3.0**(0.5)	15/15
CMA-MSR	3.2(4)	3.1(2)	6.6(2)	7.3(3)	3.5(0.5)	15/15
CMA-TPA	**2.8**(7)	3.6(0.8)	4.4(2)	5.7(4)	**3.2**(0.4)	15/15
GP1-CMAES	**1.4**(0.9)	**2.4**(3)	4.5(3)	**2.3**(2)	5.1(4)	11/15
GP5-CMAES	**1.5**(1)	**2.5**(2)	**2.9**(2)	**1.9**(1)	**2.3**(5)	14/15
IPOPCMAv3p	**2.5**(4)	**2.9**(2)	5.1(3)	5.3(6)	∞ *1258*	0/15
LHD-10xDef	**1.7**(2)	3.1(3)	6.1(6)	5.2(9)	∞ *250*	0/15
LHD-2xDefa	**1.5**(1)	**3.0**(2)	4.1(3)	**3.5**(4)	∞ *250*	0/15
RAND-2xDef	**1.6**(0.6)	3.6(3)	5.8(4)	8.7(7)	∞ *250*	0/15
RF1-CMAES	**2.2**(1)	**2.6**(2)	3.9(2)	5.1(12)	∞ *1258*	0/15
RF5-CMAES	**2.2**(3)	**2.3**(3)	**3.0**(2)	11(11)	∞ *1260*	0/15
Sifeg	**2.4**(2)	**2.3**(2)	**1.8**(0.7)	11(10)	∞ *2e4*	0/15
Sif	**2.4**(2)	**2.3**(2)	**1.8**(0.5)	24(30)	∞ *2e4*	0/15
Srr	**2.4**(2)	**2.3**(1)	**1.7**(0.8)	23(0.2)	∞ *2e4*	0/15

Table 13: 05-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_{12} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f12	*4.0e+7*:3.6	*1.6e+7*:7.6	*4.0e+6*:19	*1.6e+4*:52	*1.0e+0*:268	15/15
BSifeg	**1.3**(1)	**2.0**(1)	**1.2**(1)	8.1(5)	144(213)	6/15
BSif	**1.3**(1)	**2.1**(1)	**1.4**(0.2)	11(9)	157(334)	6/15
BSqi	**1.3**(2)	**2.0**(0.8)	**1.1**(0.4)	6.7(15)	42(34)	12/15
BSrr	**1.3**(2)	**2.0**(1)	**1.2**(0.9)	14(8)	51(63)	10/15
CMA-CSA	3.5(3)	**2.9**(3)	**2.4**(2)	4.7(1)	7.1(3)	15/15
CMA-MSR	**2.1**(6)	3.1(4)	**2.9**(2)	5.7(1)	**5.4**(6)	15/15
CMA-TPA	**2.8**(2)	3.3(2)	4.0(4)	4.6(1)	**6.1**(1)	15/15
GP1-CMAES	**2.0**(2)	**2.5**(2)	**1.9**(0.6)	**3.7**(2)	**6.2**(6)	8/15
GP5-CMAES	3.1(2)	**2.9**(2)	**3.0**(2)	21(17)	8.5(15)	6/15
IPOPCMAv3p	**1.9**(3)	**2.9**(2)	3.0(3)	4.6(1.0)	10(12)	6/15
LHD-10xDef	**1.1**(1)	**2.8**(3)	4.5(2)	8.2(4)	∞ *250*	0/15
LHD-2xDefa	**1.5**(1)	**1.9**(1)	**1.4**(0.9)	**3.0**(1)	∞ *250*	0/15
RAND-2xDef	**1.4**(3)	**2.0**(2)	**1.5**(0.4)	**2.9**(0.8)	∞ *250*	0/15
RF1-CMAES	**2.1**(1)	3.2(3)	3.2(1)	5.1(4)	22(12)	3/15
RF5-CMAES	**1.9**(2)	**1.7**(2)	5.6(2)	117(171)	∞ *1260*	0/15
Sifeg	**1.3**(1)	8.2(48)	5.1(0.4)	10(17)	50(69)	5/15
Sif	**1.3**(0.9)	4.4(1)	**2.0**(0.4)	7.5(8)	56(58)	5/15
Srr	**1.3**(1)	6.9(20)	4.6(13)	11(22)	21(17)	9/15

Table 14: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_{13} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f13	*1.0e+3*:2.8	*6.3e+2*:8.4	*4.0e+2*:17	*6.3e+1*:52	*6.3e-2*:264	15/15
BSifeg	**1.5**(2)	**1.5**(0.9)	**1.1**(0.6)	223(318)	2385(2219)	1/15
BSif	**1.5**(3)	**1.5**(0.9)	**1.1**(0.7)	227(323)	∞ *4e4*	0/15
BSqi	**1.5**(0.7)	**1.5**(0.9)	**1.1**(0.7)	208(285)	∞ *4e4*	0/15
BSrr	**1.5**(3)	**1.5**(0.5)	**1.1**(1.0)	221(422)	1122(1443)	2/15
CMA-CSA	3.2(4)	3.1(3)	**3.0**(2)	3.9(3)	**4.0**(1)	15/15
CMA-MSR	3.8(3)	**2.5**(2)	3.5(1)	4.5(2)	**3.8**(0.5)	15/15
CMA-TPA	4.6(3)	**2.9**(3)	3.5(2)	3.8(2)	**4.3**(2)	15/15
GP1-CMAES	**2.3**(2)	**1.6**(1)	**2.0**(1)	**2.3**(0.7)	70(92)	1/15
GP5-CMAES	3.2(3)	**2.1**(2)	**1.6**(0.5)	**1.4**(0.3)	11(14)	5/15
IPOPCMAv3p	3.0(4)	**2.2**(3)	3.0(2)	3.9(0.6)	10(15)	7/15
LHD-10xDef	**2.6**(1)	**3.0**(3)	4.8(3)	**2.9**(0.5)	∞ *250*	0/15
LHD-2xDefa	**2.1**(1)	**2.0**(1)	**1.9**(0.3)	**1.8**(0.4)	∞ *250*	0/15
RAND-2xDef	**2.0**(1)	**1.8**(1)	**1.8**(0.9)	**1.7**(0.5)	∞ *250*	0/15
RF1-CMAES	3.2(2)	**2.0**(1)	**2.4**(1)	8.1(19)	69(82)	1/15
RF5-CMAES	3.5(2)	**2.2**(2)	**2.1**(1)	35(21)	∞ *1252*	0/15
Sifeg	**1.5**(3)	**1.5**(0.8)	**1.1**(0.8)	83(183)	2257(3322)	1/15
Sif	**1.5**(3)	**1.5**(0.9)	**1.1**(0.8)	118(282)	∞ *4e4*	0/15
Srr	**1.5**(1)	**1.5**(1.0)	**1.1**(0.7)	178(277)	1104(1218)	2/15

Table 15: 05-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{14} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f14	*1.6e+1*:3.0	*1.0e+1*:10	*6.3e+0*:15	*2.5e-1*:53	*1.0e-5*:251	15/15
BSifeg	**3.0**(2)	**1.5**(0.7)	**1.4**(0.6)	8.7(7)	∞ *5e4*	0/15
BSif	**3.0**(3)	**1.5**(1.0)	**1.4**(1)	10(11)	∞ *5e4*	0/15
BSqi	**3.0**(3)	**1.5**(0.9)	**1.4**(0.8)	5.6(4)	∞ *5e4*	0/15
BSrr	**3.0**(3)	**1.5**(0.8)	**1.4**(1)	7.7(9)	∞ *5e4*	0/15
CMA-CSA	4.1(4)	**1.7**(3)	**2.6**(2)	3.2(1)	**3.9**(0.6)	15/15
CMA-MSR	4.3(4)	**2.5**(1)	**2.5**(1)	4.2(0.9)	**4.1**(0.6)	15/15
CMA-TPA	3.5(10)	**2.1**(1)	**2.7**(4)	3.6(1)	**4.0**(0.7)	15/15
GP1-CMAES	3.2(2)	**1.6**(1.0)	**1.9**(2)	**2.2**(0.9)	∞ *1258*	0/15
GP5-CMAES	4.0(4)	**1.8**(1)	**1.5**(2)	**1.6**(0.6)	∞ *1260*	0/15
IPOPCMAv3p	4.8(3)	**2.4**(3)	**2.4**(2)	3.8(0.5)	24(23)	3/15
LHD-10xDef	**2.0**(1)	**1.2**(1)	**2.2**(2)	3.2(0.4)	∞ *250*	0/15
LHD-2xDefa	**2.2**(3)	**1.5**(1)	**1.4**(0.6)	**2.0**(0.9)	∞ *250*	0/15
RAND-2xDef	3.0(2)	**1.4**(2)	**1.7**(0.9)	4.8(10)	∞ *250*	0/15
RF1-CMAES	3.2(5)	**2.1**(2)	**2.8**(4)	4.2(1)	∞ *1258*	0/15
RF5-CMAES	**2.8**(2)	**1.2**(1)	**2.0**(3)	81(163)	∞ *1260*	0/15
Sifeg	**3.0**(3)	**1.5**(1)	**1.4**(0.7)	**2.7**(3)	∞ *5e4*	0/15
Sif	**3.0**(3)	**1.5**(1)	**1.4**(0.7)	**2.8**(1)	∞ *5e4*	0/15
Srr	**3.0**(3)	**1.5**(1)	**1.4**(0.9)	**1.9**(1)	∞ *5e4*	0/15

Table 16: 05-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{15} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f15	*1.6e+2*:3.0	*1.0e+2*:13	*6.3e+1*:24	*4.0e+1*:55	*1.6e+1*:289	5/5
BSifeg	**2.9**(2)	3.1(8)	10(34)	70(255)	160(181)	9/15
BSif	**2.9**(3)	**1.8**(1)	27(7)	50(220)	228(268)	7/15
BSqi	**2.9**(2)	**2.7**(12)	26(80)	52(43)	184(139)	8/15
BSrr	**2.9**(3)	**2.6**(6)	10(14)	104(142)	237(276)	7/15
CMA-CSA	4.5(5)	**2.3**(1.0)	**2.6**(1.0)	**2.0**(1.0)	**1.4**(0.9)	15/15
CMA-MSR	5.1(6)	**2.3**(3)	**2.4**(2)	**2.0**(1)	**1.0**(0.5)	15/15
CMA-TPA	7.8(14)	**2.8**(3)	**2.5**(1)	**2.2**(0.8)	**1.2**(0.7)	15/15
GP1-CMAES	4.6(3)	**1.9**(1)	**1.9**(1)	**1.4**(0.5)	**1.2**(1)	15/15
GP5-CMAES	3.5(3)	**1.6**(1)	**1.6**(1)	**1.1**(0.4)	3.0(4)	11/15
IPOPCMAv3p	**2.3**(2)	**1.5**(0.6)	**1.6**(1.0)	**1.8**(1)	**1.3**(0.8)	15/15
LHD-10xDef	4.4(5)	**2.9**(3)	3.8(2)	**2.7**(0.5)	**1.3**(2)	9/15
LHD-2xDefa	**2.1**(1)	**1.5**(0.8)	**1.5**(0.5)	**1.5**(0.4)	**1.0**(1)	9/15
RAND-2xDef	3.7(3)	**1.6**(1)	**1.3**(0.6)	**1.4**(1.0)	**1.2**(2)	8/15
RF1-CMAES	5.4(4)	**2.2**(1)	**2.4**(1)	**1.7**(1)	**1.3**(2)	14/15
RF5-CMAES	3.5(5)	**1.6**(0.8)	**2.4**(3)	7.0(9)	5.7(8)	8/15
Sifeg	**2.9**(3)	**1.2**(1.0)	**1.0**(0.4)	**0.90**(0.4)	52(63)	14/15
Sif	**2.9**(3)	**1.2**(0.7)	**1.0**(0.3)	**0.87**(0.7)	51(78)	13/15
Srr	**2.9**(2)	**1.2**(0.9)	**1.0**(0.2)	**0.85**(0.6)	39(56)	14/15

Table 17: 05-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{16} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f16	*4.0e+1*:4.8	*2.5e+1*:16	*1.6e+1*:46	*1.0e+1*:120	*4.0e+0*:334	15/15
BSifeg	**1.7**(2)	**1.4**(0.9)	**0.85**(0.8)	**1.3**(2)	40(46)	15/15
BSif	**1.7**(1)	**1.4**(1.0)	**1.0**(1)	10(35)	32(25)	15/15
BSqi	**1.7**(2)	**1.4**(0.7)	**1.1**(0.4)	**2.5**(6)	22(37)	15/15
BSrr	**1.7**(2)	**1.4**(0.7)	**1.1**(1)	**1.4**(0.8)	30(21)	15/15
CMA-CSA	3.7(4)	**2.9**(5)	3.1(3)	**2.2**(1)	**1.5**(0.7)	15/15
CMA-MSR	**2.0**(1)	**2.2**(2)	**1.9**(2)	5.9(11)	4.6(9)	15/15
CMA-TPA	**2.3**(1)	**2.9**(3)	**3.0**(2)	**1.7**(2)	**1.8**(0.6)	15/15
GP1-CMAES	**1.6**(1)	**1.3**(1)	**0.90**(0.8)	**1.2**(0.9)	**1.4**(2)	13/15
GP5-CMAES	**2.0**(2)	**1.5**(0.4)	**2.7**(4)	**1.3**(2)	**1.9**(2)	13/15
IPOPCMAv3p	**1.4**(0.6)	**1.2**(0.9)	**1.5**(1)	**2.4**(2)	**2.9**(3)	11/15
LHD-10xDef	**1.4**(0.7)	**1.6**(1)	**1.9**(2)	**1.5**(0.5)	**1.4**(0.8)	7/15
LHD-2xDefa	**2.5**(2)	**1.4**(0.6)	**1.6**(1)	**2.2**(2)	11(19)	1/15
RAND-2xDef	**1.3**(0.6)	**1.2**(1)	**1.8**(2)	**1.7**(3)	**1.3**(1)	7/15
RF1-CMAES	**1.3**(1)	**0.90**(1)	**1.3**(2)	**1.8**(2)	**2.1**(2)	11/15
RF5-CMAES	**2.5**(3)	**1.3**(1)	**1.1**(1)	**1.7**(5)	3.2(5)	9/15
Sifeg	**1.7**(1)	**1.6**(2)	**0.89**(0.7)	**0.62**(0.4)	7.2(6)	15/15
Sif	**1.7**(1)	**1.6**(1)	**0.89**(0.4)	**0.69**(0.5)	7.7(6)	15/15
Srr	**1.7**(1)	**1.6**(1)	**0.88**(0.5)	**0.68**(0.6)	3.4(10)	15/15

Table 18: 05-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{17} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f17	*1.0e+1*:5.2	*6.3e+0*:26	*4.0e+0*:57	*2.5e+0*:110	*6.3e-1*:412	15/15
BSifeg	6.3(21)	5.5(14)	174(222)	189(123)	268(547)	5/15
BSif	7.0(3)	298(478)	322(447)	188(116)	157(92)	7/15
BSqi	4.1(4)	139(944)	141(444)	160(229)	150(162)	7/15
BSrr	4.3(3)	6.4(7)	182(447)	137(183)	229(260)	6/15
CMA-CSA	4.2(5)	**1.8**(1)	**1.5**(1.0)	**1.0**(0.4)	**0.61**(0.3)	15/15
CMA-MSR	4.2(2)	**2.0**(2)	**1.5**(0.6)	**1.1**(0.5)	**0.60**(0.1)	15/15
CMA-TPA	24(72)	6.0(3)	3.5(1)	**2.1**(0.7)	**1.5**(3)	15/15
GP1-CMAES	4.5(4)	**1.6**(1)	**1.1**(0.6)	**0.78**(0.1)	**0.45**(0.3)	15/15
GP5-CMAES	3.6(6)	**1.6**(2)	**1.6**(4)	**1.8**(0.1)	**2.5**(3)	10/15
IPOPCMAv3p	4.1(4)	**1.8**(0.9)	**1.3**(0.9)	**1.2**(0.5)	**0.81**(0.9)	14/15
LHD-10xDef	**2.1**(1)	**2.4**(2)	**1.8**(1)	**1.4**(0.4)	4.4(5)	2/15
LHD-2xDefa	**2.4**(2)	**1.4**(0.6)	**1.0**(0.7)	**1.1**(2)	8.9(13)	1/15
RAND-2xDef	**2.3**(3)	**1.2**(1)	**1.4**(0.8)	**1.8**(2)	4.3(4)	2/15
RF1-CMAES	3.0(3)	**2.2**(2)	3.3(7)	3.9(11)	4.1(5)	7/15
RF5-CMAES	4.8(7)	8.6(2)	9.0(11)	10(8)	22(8)	2/15
Sifeg	3.9(2)	10(68)	76(84)	61(206)	109(171)	9/15
Sif	3.9(3)	**1.4**(1)	132(428)	112(142)	112(80)	9/15
Srr	3.9(3)	**1.5**(1)	99(343)	121(360)	191(303)	6/15

Table 19: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_{18} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f18	*6.3e+1*:3.4	*4.0e+1*:7.2	*2.5e+1*:20	*1.6e+1*:58	*1.6e+0*:318	15/15
BSifeg	**1.5**(3)	**2.2**(2)	75(529)	114(282)	184(219)	8/15
BSif	**1.5**(0.9)	**2.2**(3)	115(414)	120(119)	228(210)	7/15
BSqi	**1.5**(2)	**2.3**(3)	91(3)	90(299)	302(205)	6/15
BSrr	**1.5**(3)	**2.3**(2)	8.4(26)	106(168)	204(290)	7/15
CMA-CSA	4.4(6)	3.9(1)	**2.1**(1)	**1.4**(0.7)	**2.1**(2)	15/15
CMA-MSR	**1.6**(1.0)	**2.0**(2)	**2.0**(1)	**1.3**(0.7)	4.1(12)	15/15
CMA-TPA	**2.5**(1)	**2.7**(3)	**1.8**(2)	**1.1**(0.8)	**1.4**(2)	15/15
GP1-CMAES	**2.4**(4)	3.6(2)	**2.1**(2)	**1.2**(0.5)	**1.2**(1)	14/15
GP5-CMAES	**1.7**(0.9)	**3.0**(3)	5.2(11)	**2.1**(3)	5.8(6)	7/15
IPOPCMAv3p	**1.2**(0.7)	**1.2**(2)	**1.8**(2)	**1.3**(1)	**1.4**(0.3)	14/15
LHD-10xDef	**2.6**(5)	3.4(4)	3.3(3)	**2.0**(0.8)	12(12)	1/15
LHD-2xDefa	**1.3**(2)	**1.6**(2)	**2.2**(2)	**1.3**(0.6)	5.5(7)	2/15
RAND-2xDef	**1.4**(1)	**2.2**(2)	**1.7**(0.9)	**1.5**(1)	12(14)	1/15
RF1-CMAES	**1.5**(3)	**1.7**(2)	**1.3**(1)	**0.89**(0.9)	**2.2**(3)	11/15
RF5-CMAES	**2.2**(2)	**3.0**(4)	9.5(22)	7.6(8)	29(28)	2/15
Sifeg	**1.5**(3)	**2.1**(1)	6.1(0.4)	29(19)	145(192)	9/15
Sif	**1.5**(2)	**2.1**(2)	63(232)	40(145)	118(135)	10/15
Srr	**1.5**(3)	**2.2**(1)	3.6(10)	38(50)	67(191)	11/15

Table 20: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_{19} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f19	*1.6e-1*:172	*1.0e-1*:242	*6.3e-2*:675	*4.0e-2*:3078	*2.5e-2*:4946	15/15
BSifeg	917(1787)	909(830)	1077(864)	∞	∞ *5e4*	0/15
BSif	732(937)	694(668)	∞	∞	∞ *5e4*	0/15
BSqi	969(780)	1440(975)	∞	∞	∞ *5e4*	0/15
BSrr	737(913)	925(869)	∞	∞	∞ *5e4*	0/15
CMA-CSA	**154**(83)	**153**(146)	**70**(55)	**19**(26)	**15**(16)	15/15
CMA-MSR	**222**(95)	**306**(76)	**229**(559)	**139**(167)	**137**(100)	7/15
CMA-TPA	**91**(67)	**84**(68)	**39**(44)	**18**(18)	**14**(13)	15/15
GP1-CMAES	∞	∞	∞	∞	∞ *1260*	0/15
GP5-CMAES	∞	∞	∞	∞	∞ *1262*	0/15
IPOPCMAv3p	∞	∞	∞	∞	∞ *1258*	0/15
LHD-10xDef	∞	∞	∞	∞	∞ *250*	0/15
LHD-2xDefa	∞	∞	∞	∞	∞ *250*	0/15
RAND-2xDef	∞	∞	∞	∞	∞ *250*	0/15
RF1-CMAES	∞	∞	∞	∞	∞ *1258*	0/15
RF5-CMAES	∞	∞	∞	∞	∞ *1262*	0/15
Sifeg	437(515)	477(723)	1039(1787)	∞	∞ *5e4*	0/15
Sif	521(333)	1385(2297)	1021(1558)	∞	∞ *5e4*	0/15
Srr	670(770)	671(1338)	1042(946)	∞	∞ *5e4*	0/15

Table 21: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_{20} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f20	*6.3e+3*:5.1	*4.0e+3*:8.4	*4.0e+1*:15	*2.5e+0*:69	*1.0e+0*:851	15/15
BSifeg	3.0(2)	**2.0**(1)	**1.9**(1)	**2.2**(3)	9.3(20)	14/15
BSif	3.0(2)	**2.0**(1)	**2.0**(0.5)	**2.1**(2)	23(35)	12/15
BSqi	3.0(1)	**2.0**(0.7)	**1.8**(1)	**2.3**(1)	8.7(27)	15/15
BSrr	3.0(2)	**2.0**(0.9)	**1.8**(2)	**1.5**(2)	11(30)	14/15
CMA-CSA	**2.3**(3)	**2.0**(1)	3.7(0.9)	**2.5**(0.9)	9.2(5)	15/15
CMA-MSR	3.3(5)	**2.4**(3)	4.9(2)	**2.8**(0.8)	1666(1444)	4/15
CMA-TPA	**2.8**(3)	**2.3**(2)	3.8(1)	3.1(2)	17(19)	15/15
GP1-CMAES	**1.9**(2)	**1.8**(1)	**3.0**(2)	4.2(2)	11(4)	2/15
GP5-CMAES	**2.5**(2)	**1.8**(2)	**2.1**(0.9)	**2.1**(1.0)	∞ *1260*	0/15
IPOPCMAv3p	**2.1**(3)	**1.8**(1)	4.1(2)	3.3(2)	21(21)	1/15
LHD-10xDef	**1.8**(2)	**1.4**(0.8)	6.3(4)	11(8)	∞ *250*	0/15
LHD-2xDefa	**2.8**(2)	**2.2**(1)	**2.2**(0.8)	7.6(8)	∞ *250*	0/15
RAND-2xDef	**2.0**(2)	**2.1**(2)	**2.6**(1.0)	3.3(2)	∞ *250*	0/15
RF1-CMAES	**2.9**(5)	**2.4**(3)	3.9(2)	5.6(9)	∞ *1258*	0/15
RF5-CMAES	**2.5**(2)	**1.8**(1)	25(26)	20(24)	∞ *1260*	0/15
Sifeg	3.0(2)	**2.1**(0.3)	**1.8**(0.6)	**1.1**(0.4)	**3.1**(0.6)	15/15
Sif	3.0(2)	**2.1**(1)	**1.8**(0.4)	**1.2**(0.3)	**6.6**(21)	14/15
Srr	3.0(2)	**2.1**(0.9)	**1.8**(1)	**0.83**(0.5)	**2.8**(6)	15/15

Table 22: 05-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_{21} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f21	*4.0e+1*:3.9	*2.5e+1*:11	*1.6e+1*:31	*6.3e+0*:73	*1.6e+0*:347	5/5
BSifeg	**2.9**(3)	**1.6**(2)	11(0.4)	90(245)	141(232)	8/15
BSif	**2.9**(2)	**1.7**(1)	101(4)	183(253)	290(217)	5/15
BSqi	**2.9**(2)	**2.4**(0.9)	16(58)	85(226)	221(325)	6/15
BSrr	**2.9**(3)	**1.6**(2)	14(50)	73(23)	137(142)	9/15
CMA-CSA	**2.5**(2)	**1.9**(1)	**1.7**(2)	**1.8**(1)	92(2)	14/15
CMA-MSR	3.2(6)	**2.0**(0.8)	**1.7**(1)	3.8(2)	249(589)	13/15
CMA-TPA	**2.4**(2)	**2.1**(2)	**2.0**(0.9)	**2.0**(1.0)	45(148)	15/15
GP1-CMAES	**1.1**(1)	**1.4**(2)	**0.99**(0.5)	4.7(22)	4.9(6)	7/15
GP5-CMAES	**2.5**(2)	**1.7**(1)	**0.89**(0.4)	**1.4**(3)	4.7(10)	8/15
IPOPCMAv3p	**1.8**(2)	**1.6**(1)	**1.7**(2)	10(18)	15(20)	3/15
LHD-10xDef	**1.5**(1)	**1.7**(2)	**1.5**(2)	**2.2**(0.6)	**2.0**(3)	5/15
LHD-2xDefa	**1.7**(2)	**2.1**(2)	**1.3**(0.7)	**1.5**(0.9)	**1.7**(1)	5/15
RAND-2xDef	**1.4**(2)	**1.1**(1)	**1.3**(1.0)	**1.4**(0.9)	**3.3**(3)	3/15
RF1-CMAES	**2.0**(2)	**1.8**(0.9)	**1.4**(2)	4.1(5)	4.8(3)	7/15
RF5-CMAES	**1.8**(2)	**2.6**(2)	**2.9**(7)	7.5(8)	10(5)	5/15
Sifeg	**2.9**(2)	**1.4**(1)	**0.82**(0.6)	110(0.4)	195(505)	7/15
Sif	**2.9**(2)	**1.6**(1)	**0.93**(0.5)	158(341)	150(91)	8/15
Srr	**2.9**(2)	**1.4**(1)	**0.84**(0.5)	73(195)	179(411)	7/15

Table 23: 05-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_{22} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f22	*6.3e+1*:3.6	*4.0e+1*:15	*2.5e+1*:32	*1.0e+1*:71	*1.6e+0*:341	5/5
BSifeg	3.5(2)	7.4(46)	15(80)	34(14)	88(126)	10/15
BSif	3.5(4)	5.0(0.6)	114(1)	102(180)	217(183)	7/15
BSqi	3.5(4)	3.8(10)	12(9)	31(43)	59(178)	13/15
BSrr	3.5(2)	5.5(16)	13(0.4)	37(178)	100(155)	10/15
CMA-CSA	**2.7**(2)	**1.6**(1)	**1.4**(0.9)	4.1(6)	151(274)	14/15
CMA-MSR	4.3(6)	**1.8**(2)	5.2(15)	14(16)	113(136)	15/15
CMA-TPA	3.7(3)	**1.9**(1)	**1.2**(0.6)	**2.5**(0.7)	253(236)	12/15
GP1-CMAES	**2.6**(3)	**1.4**(1)	**1.2**(0.9)	3.6(0.1)	10(10)	4/15
GP5-CMAES	**2.9**(2)	7.5(7)	4.6(3)	4.3(6)	11(17)	4/15
IPOPCMAv3p	**2.8**(1)	**2.2**(3)	**1.7**(1)	5.8(9)	6.3(9)	6/15
LHD-10xDef	**2.2**(2)	**2.0**(0.8)	**2.3**(2)	**1.9**(0.5)	**2.5**(2)	4/15
LHD-2xDefa	3.3(3)	**1.8**(2)	**1.3**(0.8)	**1.4**(2)	**1.8**(3)	5/15
RAND-2xDef	**1.8**(2)	**1.5**(0.9)	**0.98**(0.6)	**0.79**(0.6)	**2.4**(6)	4/15
RF1-CMAES	3.1(5)	**2.0**(2)	**1.3**(1)	5.5(13)	3.2(5)	9/15
RF5-CMAES	**2.1**(4)	4.7(27)	**2.9**(2)	7.3(3)	24(16)	2/15
Sifeg	3.4(3)	**1.8**(2)	115(0.7)	123(212)	116(262)	9/15
Sif	3.4(3)	**1.7**(1)	5.8(18)	117(379)	125(83)	9/15
Srr	3.4(2)	**1.6**(1)	**2.5**(0.8)	62(177)	75(98)	11/15

Table 24: 05-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_{23} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f23	*1.0e+1*:3.0	*6.3e+0*:9.0	*4.0e+0*:33	*2.5e+0*:84	*1.0e+0*:518	15/15
BSifeg	**2.6**(2)	**2.5**(2)	**1.5**(0.8)	**1.7**(1.0)	4.3(5)	15/15
BSif	**2.6**(3)	**2.5**(1)	**1.6**(2)	**1.9**(3)	3.3(5)	15/15
BSqi	**2.6**(2)	**2.4**(2)	**1.9**(3)	**2.0**(2)	6.6(6)	15/15
BSrr	**2.6**(2)	**2.5**(2)	**1.6**(1)	**2.0**(2)	3.7(6)	15/15
CMA-CSA	**2.3**(2)	3.2(3)	5.9(5)	8.9(8)	13(18)	15/15
CMA-MSR	**2.5**(4)	3.7(10)	3.1(3)	6.0(3)	3.2(4)	15/15
CMA-TPA	3.2(2)	3.8(5)	3.4(3)	12(5)	16(12)	15/15
GP1-CMAES	**1.9**(3)	**2.7**(2)	3.2(3)	6.5(4)	4.9(2)	6/15
GP5-CMAES	**2.4**(4)	**2.0**(2)	**2.4**(2)	3.0(4)	**2.2**(4)	11/15
IPOPCMAv3p	**2.3**(4)	**2.2**(3)	3.4(1)	3.8(3)	12(29)	3/15
LHD-10xDef	3.9(5)	3.4(3)	3.3(2)	4.3(8)	6.8(5)	1/15
LHD-2xDefa	3.1(4)	3.8(2)	3.8(3)	10(11)	∞ *250*	0/15
RAND-2xDef	**2.5**(1)	**2.7**(2)	**2.1**(2)	4.6(4)	7.1(7)	1/15
RF1-CMAES	**1.8**(2)	**2.0**(2)	**3.0**(2)	8.7(13)	∞ *1260*	0/15
RF5-CMAES	**2.4**(2)	**1.5**(3)	3.3(2)	4.0(3)	∞ *1288*	0/15
Sifeg	3.4(9)	**2.9**(3)	**1.7**(1)	**2.8**(2)	**2.7**(2)	15/15
Sif	3.4(2)	3.2(3)	**1.8**(1)	3.0(3)	**2.8**(3)	15/15
Srr	3.4(2)	3.2(3)	**2.3**(3)	3.4(3)	**2.5**(3)	15/15

Table 25: 05-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_{24} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f24	*6.3e+1*:15	*4.0e+1*:37	*4.0e+1*:37	*2.5e+1*:118	*1.6e+1*:692	15/15
BSifeg	**1.6**(0.4)	3.8(2)	3.8(6)	32(14)	20(19)	14/15
BSif	**1.6**(0.9)	**2.5**(7)	**2.5**(4)	37(52)	17(11)	14/15
BSqi	**2.8**(0.3)	**2.9**(3)	**2.9**(7)	31(3)	18(18)	14/15
BSrr	**1.5**(0.6)	3.3(2)	3.3(0.6)	41(100)	32(77)	11/15
CMA-CSA	**2.0**(1)	**2.1**(2)	**2.1**(0.9)	**1.5**(2)	**1.6**(2)	15/15
CMA-MSR	**2.2**(2)	**2.6**(0.6)	**2.6**(1)	**1.8**(0.8)	**0.93**(2)	15/15
CMA-TPA	**1.9**(1)	**1.9**(0.8)	**1.9**(1)	**1.9**(1.0)	**1.4**(1)	15/15
GP1-CMAES	**1.7**(1)	**1.7**(0.5)	**1.7**(0.9)	**1.3**(1.0)	**1.1**(1.0)	13/15
GP5-CMAES	**1.5**(1.0)	**1.1**(0.4)	**1.1**(0.4)	**2.5**(2)	**1.6**(2)	9/15
IPOPCMAv3p	**2.0**(1)	**2.3**(2)	**2.3**(1)	**1.6**(1)	**1.1**(0.7)	12/15
LHD-10xDef	**2.7**(2)	6.5(2)	6.5(9)	4.8(3)	∞ *250*	0/15
LHD-2xDefa	**1.9**(2)	3.4(4)	3.4(5)	∞	∞ *250*	0/15
RAND-2xDef	**1.7**(2)	7.3(9)	7.3(6)	15(16)	∞ *250*	0/15
RF1-CMAES	**1.9**(1)	**2.4**(3)	**2.4**(3)	**2.4**(4)	**1.9**(2)	9/15
RF5-CMAES	**1.8**(2)	3.6(6)	3.6(6)	4.1(4)	**2.9**(2)	7/15
Sifeg	**1.8**(2)	**1.7**(2)	**1.7**(2)	13(2)	4.8(2)	15/15
Sif	**1.7**(1)	**1.7**(3)	**1.7**(2)	**2.9**(2)	4.8(11)	15/15
Srr	**1.8**(0.6)	**1.6**(1)	**1.6**(1)	7.3(3)	5.8(1.0)	15/15

References

[1] Asma Atamna. Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB noiseless testbed. In Laredo et al. [8], pages 1135–1142.

[2] Anne Auger, Steffen Finck, Nikolaus Hansen, and Raymond Ros. BBOB 2009: Comparison tables of all algorithms on all noiseless functions. Technical Report RT-0383, INRIA, April 2010.

[3] Lukás Bajer, Zbynek Pitra, and Martin Holena. Benchmarking gaussian processes and random forests surrogate models on the BBOB noiseless testbed. In Laredo et al. [8], pages 1143–1150.

[4] Dimo Brockhoff, Bernd Bischl, and Tobias Wagner. The impact of initial designs on the performance of matsumoto on the noiseless BBOB-2015 testbed: A preliminary study. In Laredo et al. [8], pages 1159–1166.

[5] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009. Updated February 2010.

[6] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2012: Experimental setup. Technical report, INRIA, 2012.

[7] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009. Updated February 2010.

[8] Juan Luis Jiménez Laredo, Sara Silva, and Anna Isabel Esparcia-Alcázar, editors. *Genetic and Evolutionary Computation Conference, GECCO 2015, Madrid, Spain, July 11-15, 2015, Companion Material Proceedings*. ACM, 2015.

[9] Petr Posík and Petr Baudis. Dimension selection in axis-parallel brent-step method for black-box optimization of separable continuous functions. In Laredo et al. [8], pages 1151–1158.